Paweł Kruczkiewicz - zadanie 1

March 21, 2022

1 Badanie klątwy wymiaru

Przedmiot: Podstawy Uczenia Maszynowego Wykonał: Paweł Kruczkiewicz Kraków, 14.03.2022 r.

Celem ćwiczenia jest zbadanie problemu klątwy wymiaru, który jest jednym z trudniejszych problemów w uczeniu maszynowym. Poniżej przeanalizowane zostaną trzy sposoby na pokazanie tego problemu. Opierają się na analizie metodą Monte Carlo zbioru puntów w hiperszeŝcianie a następnie na : 1. Sprawdzeniu kąta między dwoma wylosowanymi n-wymiarowymi wektorami. 2. Wyliczeniu, jaka część lpunktów w hiperszeŝcianie znajduje się wewnątrz kuli wyznaczonej przez 2 losowo wybrane punkty 3. Wyliczeniu stosunku różnicy i średniej dla dwóch długości odcinków powstałych przez wylosowanie 3 punktów.

2 Przydatne funkcje

- `get_random_points_from_hipercube(N, D)` - **Losowanie `N` punktów w `D`-wymiarowej przestrzeni**, gdzie wartości punktów w poszczególnych wymiarach są w zakresie [0, 1)
- `calculate_random_angles(P, N)` - Zwraca **listę `N` kątów** dla `N` wylosowanych par ze zbioru pubków `P`
- `calculate_percent_of_points_in_random_sphere(P, N)` - Zwraca **procent punktów znajdujących sie w sferze wyznaczonej przez 2 losowo wybrane punkty** ze zbioru `P` w formie listy długości `N`
- `calucalate_percent_of_points(P, N)` - Zwraca % średniej dwóch odległości w stosunku do różnicy między nimi, gdzie odległością są długości dwóch odcinków z trzech losowo wylosowanych punktów
- tworzenie listy średnich i odchyleń standardowych dla danych wymiarów dla danej funkcji

Dzięki tej funkcji możemy łatwo przeanalizować wyniki powyżej zaimplementowanych funkcji. Przyjęto, że liczba powtórzeń eksperymentu będzie równa 100.

W implementacji losujemy również zbiór punktów w hiperszeŝcianie. Przyjęto, że liczba punktów w nim wynosi 10000, bo taka liczba liczy się względnie niedługo dla każdej z powyższych funkcji.

- rysowanie wykresu słupkowego
- rysowanie histogramu
- rysowanie wykresu słupkowego wraz z linią

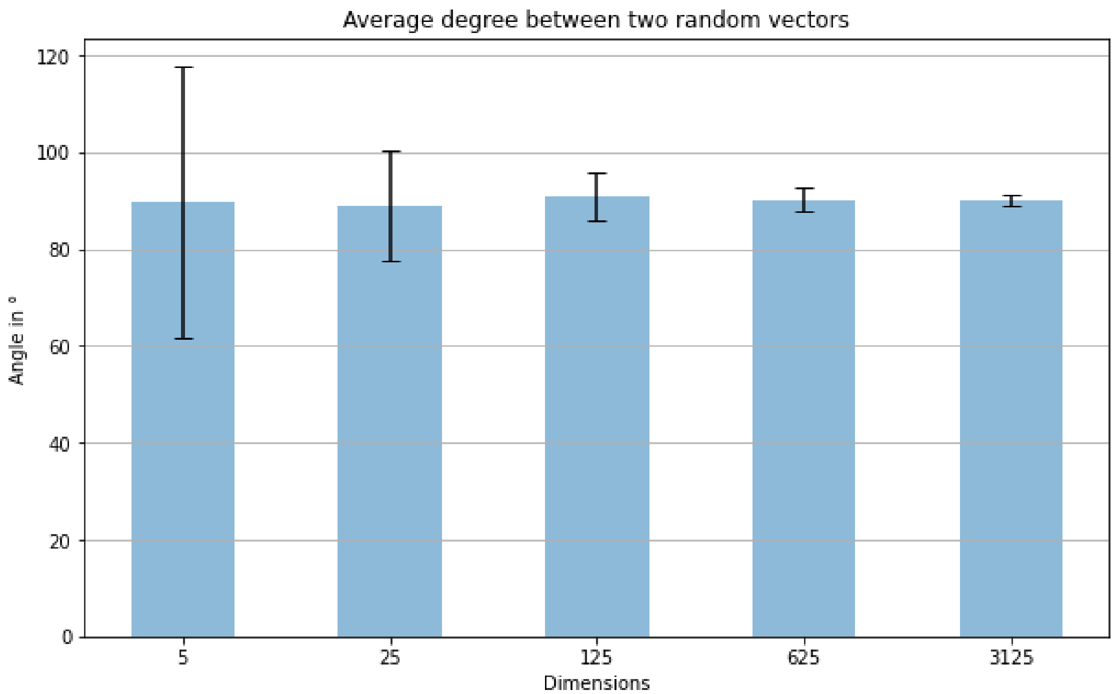
3 Wyniki

3.1 Badane wymiary

Przyjęto, że wszystkie punkty zostaną zbadane dla rosnącej potęgowo liczby wymiarów, czyi dla kolejnych potęg 5.

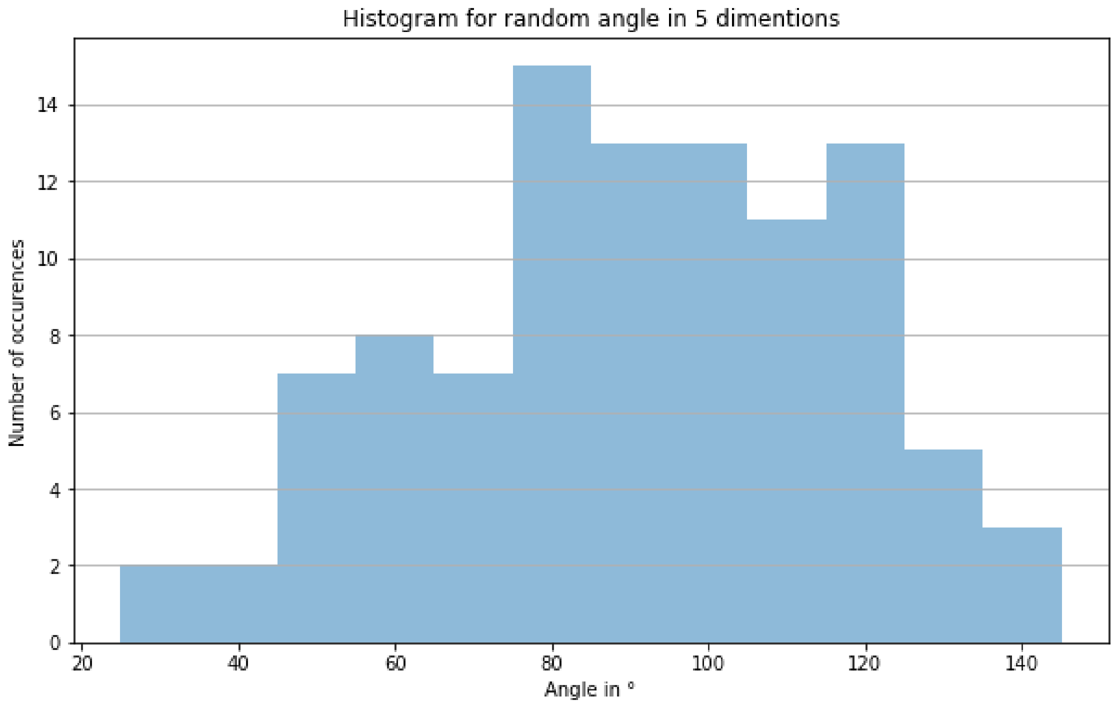
3.2 Losowe kąty

Przyjrzyjmy sie najpierw, jak wyglądają wykresy słupkowe dla kolejnych wymiarów.

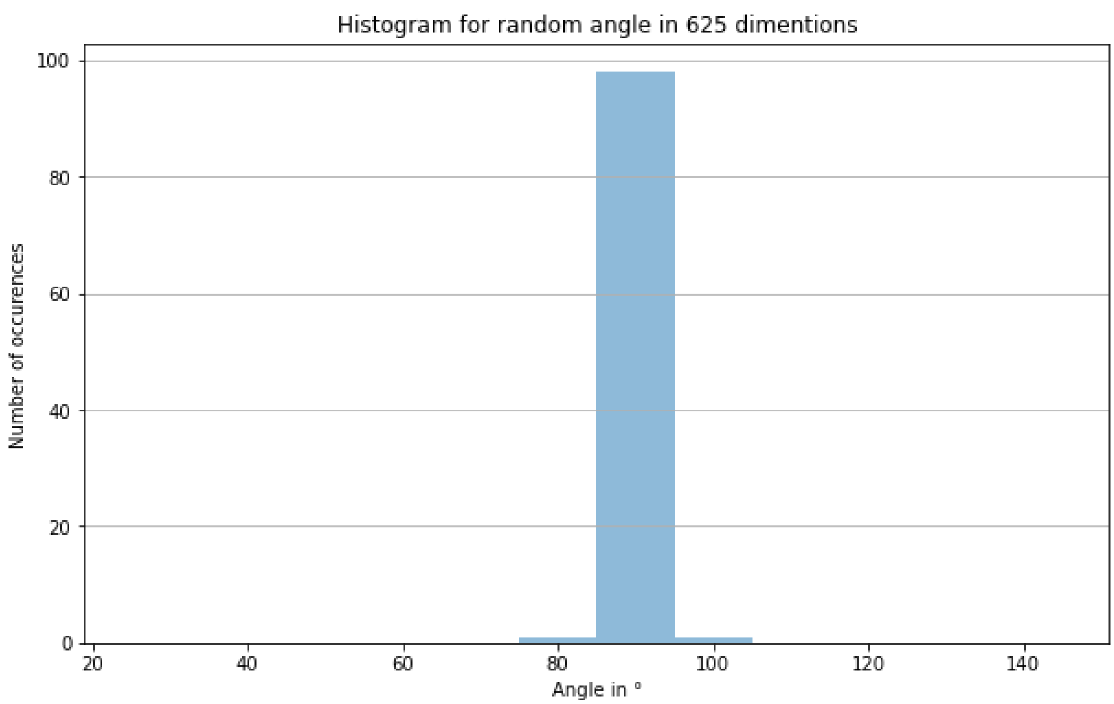


Wartości oscylują około 90 stopni. Odchylenia standardowe jednak zmniejszają się znacząco z każdym kolejnym wymiarem. Prześledźmy histogram dla dwóch skrajnie różnych wymiarów:

5 wymiarów

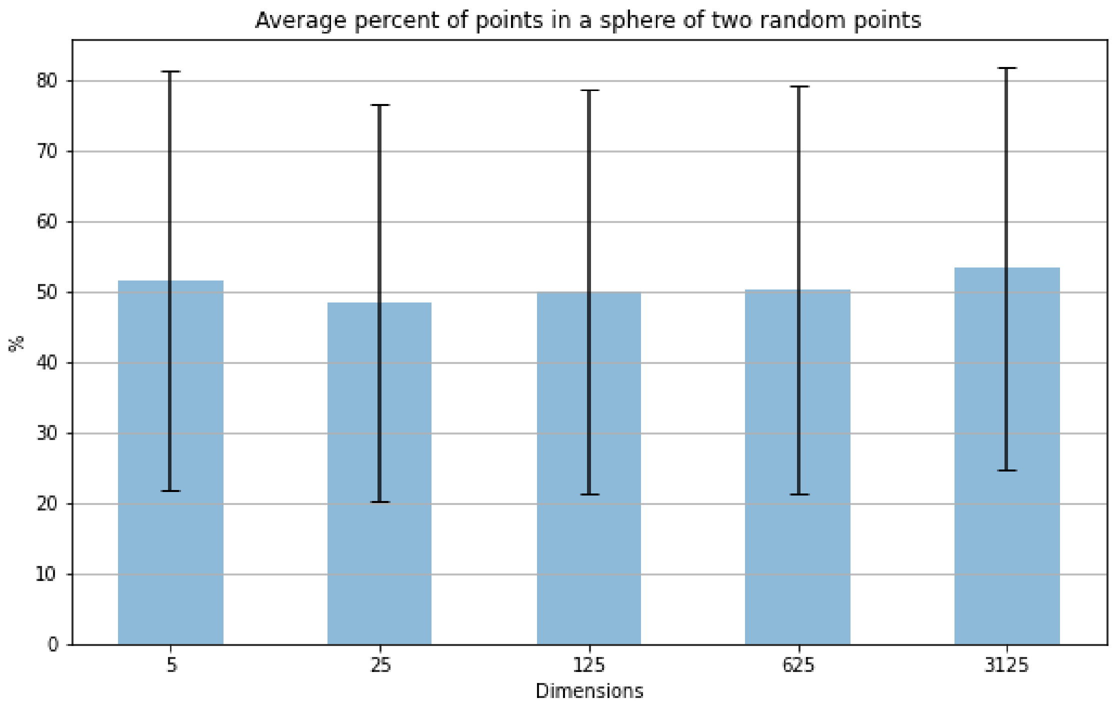


625 wymiarów



Widać wyraźna tendencję - wraz ze wrostem liczby wymiarów coraz trudniej jest wylosować wektor znacząco różny od 90 stopni. Dowodem na to może być to, że niemal wszystkie wartości dla 625 wymiarów lądują w przedziale [85, 95) stopnia.

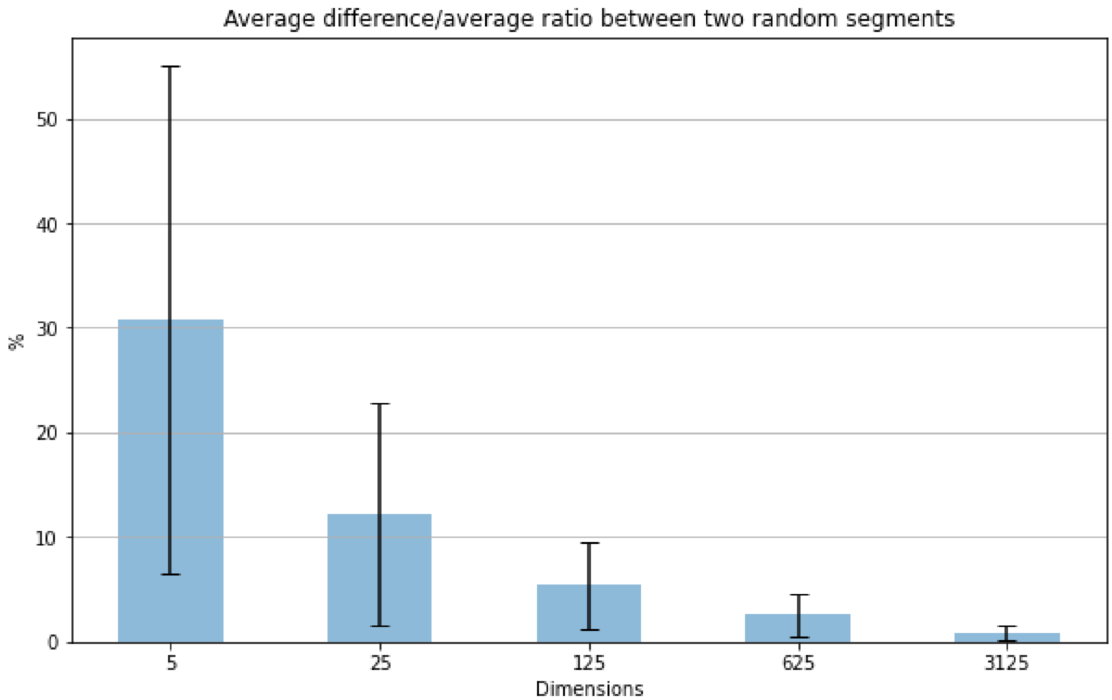
3.3 Wykres słupkowy dla kuli powstałej z losowo wybranych 2 punktów.



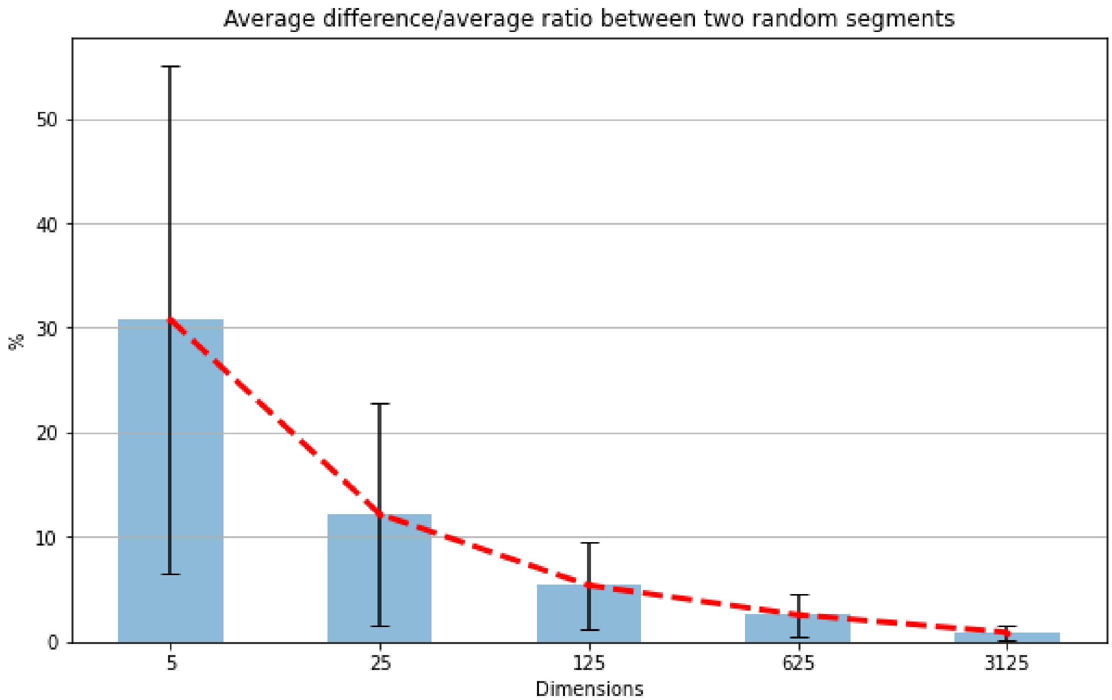
Wynik jest zaskakujący. W tym przypadku nie widzimy już tendencji do zmniejszania się odchylenia standardowego ani też mierzona wartość nie dąży do zera. Co więcej - przed eksperymentem można było przypuszczać, że słupki będą maleć ze względu na analogiczność tego eksperymentu z mierzeniem stosunku hiperkuli do opisanego na niej hipersześcianu w n-wymiarach.

Okazuje się jednak, że losowanie wymiaru kuli *w kontrolowanych warunkach* nie prowadzi do takich problemów.

3.4 Stosunek różnicy do średniej długości dwóch losowo wybranych odcinków



Jak widać na powyższym rysunku - mierzony stosunek wyraźnie maleje. Z każdym koejnym wymiarem nie tylko zanika różnorodność, ale mierzona wartość dąży do zera. Dobrze to widać na wykresie liniowym, który nałożymy na powyższy diagram.



Dzięki dużej, czerwonej, przerywanej linii wykres staje sie bardziej przekonujący.

4 Wnioski

Powyższe eksperymenty mozna podsumować słowami Witkacego:

Punkt się rozprężył w n-wymiarów przestrzeń I przestrzeń klapła Jak przekłuty balon.

Badając klątwę wymiaru dochodzimy do paradoksalnych wniosków - chcąc zwiększyć informacje o obiekcie poprzez poszerzenie wiedzy o kolejne wymiary, tracimy zdolność rozróżniania obiektów od siebie. W nieskończoności nic do niczego jest niepodobne - ta myśl jakkolwiek nasuwająca się sama, tworzy rozdarcie w ludzkim postrzeganiu, bo jak można zwiększając rozmiar informacji jednocześnie je tracić? Fakt ten mnie jako (miejmy nadzieję) przyszłego inżyniera niepokoi, a jako humanistę - przeraża. Okazuje się, że problem poznania nawet przy przyjęciu zgrubnych arystetelesowskich założeń staje się trudny do opisania nawet poprzez jedno z najbardziej racjonalnych narzędzi jakimi dysponuje człowiek, którym jest matematyka.

Pozostaje nam jedynie liczyć na to, że nasz (jak to opisuje F. Dostojewski) *umysł euklidesowy, ziemski* chociaż nie potrafi pojąć intuicyjnie tego filozoficzngo problemu, tak będzie w stanie go w jakiś sposób ujarzmić, obejść. Oczywiście wpędzamy się co zatem idzie w błędne koło - jeżeli bedziemy chcieli porównywać, które wymiary lepiej opisują przestrzeń, którą chcemy przewidzieć, dokładamy sobie kolejną (a może kolejne) metryki. Która jednak z tych metryk jest najważniejsza? To trzeba już ocenić arbitralnie, postawić granicę przejawiającej się w nas, studentach AGH bez metafizycznego pojęcia o rzeczywistości, chęci do zmierzenia i zważenia wszystkiego, co dookoła, dokonując wyboru, żegnając się z bezpiecznym siedziskiem obserwatora i przyjmując rolę rządcy czy sędziego.

Ważę ciężary o jakich nie myślał żaden cezar świata a wszystko ulata, ulata jak wata, ulata jak wata hop szklankę piwa hop